LJ2M DATASET: TOWARD BETTER UNDERSTANDING OF MUSIC LISTENING BEHAVIOR AND USER MOOD

Jen-Yu Liu, Sung-Yen Liu, and Yi-Hsuan Yang

Research Center for IT innovation, Academia Sinica, Taipei, Taiwan
{ciaua, sean51623, yang}@citi.sinica.edu.tw

ABSTRACT
Recent years have witnessed a growing interest in modeling user behaviors in multimedia research, emphasizing the need to consider human factors such as preference, activity, and emotion in system development and evaluation. Following this research line, we present in this paper the LiveJournal two-million post (LJ2M) dataset to foster research on user-centered music information retrieval. The new dataset is characterized by the great diversity of real-life listening contexts where people and music interact. It contains blog articles from the social blogging website LiveJournal, along with tags self-reporting a user’s emotional state while posting and the musical track that the user considered as the best match for the post. More importantly, the data are contributed by users spontaneously in their daily lives, instead of being collected in a controlled environment. Therefore, it offers new opportunities to understand the interrelationship among the personal, situational, and musical factors of music listening. As an example application, we present research investigating the interaction between the affective context of the listener and the affective content of music, using audio-based music emotion recognition techniques and a psycholinguistic tool. The study offers insights into the role of music in mood regulation and demonstrates how LJ2M can contribute to studies on real-world music listening behavior.

Index Terms— User-centered approach, listening context, user mood, music-listening dataset, music emotion

1. INTRODUCTION

During the past years, many researchers have called attention to the user-centered and context-aware music information researches [1–4]. For example, Schedl and Flexer proposed that users and listening context should be considered in both the development and evaluation of a music information retrieval (MIR) system [1]. However, research along this direction has been impeded by the difficulty of collecting music listening records that reflect real-world music listening behaviors. Dataset collected in a laboratory setting is usually context-deprived and therefore limited, whereas it requires great effort to collect an in-situ dataset from people’s daily life. Issues such as user burden, intrusiveness, and privacy have to be taken into account. Moreover, it would be preferable if the listening records are collected in a spontaneous way.

Another issue arises when we consider that context can be divided into two types: external and internal context [5]. According to Kokinov, context is “the set of all entities that influence system’s cognitive behavior on a particular occasion” [5]. External context includes the physical and social environment, the factors external to the user, e.g., the time and place of music listening. In contrast, internal context refers to the user’s mental state, e.g., the user’s emotional state. The external context has to be perceived by a user to change the mental state of the user and then influence the user’s behavior [5]. Internal context may exert a more direct impact on human behavior. From a research point of view, both the external and internal contexts are important. However, it is more difficult to gather information regarding the internal context.

Thanks to the prevalence of social-networking services such as Facebook and Twitter, we see new opportunities to circumvent the above issues. People are getting used to share information regarding the music they listen to through such services, where external context such as the posting time, location and social relationship can be easily recorded by the system. Moreover, as these listening records are often accompanied with an article self-reporting the user’s feeling or experience [6], the articles might provide clues to the internal context. However, even with state-of-the-art natural language processing techniques, it is still difficult to accurately estimate from an article the emotional state of the user [7–9].

In light of the above observations, we propose in this paper the LiveJournal two-million post (LJ2M) dataset harvested from the social blogging service LiveJournal to foster research on user-centered MIR. LiveJournal is unique in that, in addition to the common blogging functionality, each article is accompanied with a ‘mood’ entry and a ‘music’ entry in which the bloggers can write down their emotional states.

This work was supported by the National Science Council of Taiwan under Grant NSC 102-2221-E-001-004-MY3

1https://www.facebook.com/, https://twitter.com/
2http://www.livejournal.com/
and the musical tracks in their minds while posting, as Figure 1 shows. In this way, the mood entry and the article in combination provides rich clues to the internal context of the user, whereas the music entry connects the internal context with the short-term music preference. Moreover, as the data are contributed and shared by users spontaneously in their daily lives, many of the aforementioned concerns are mitigated.

The LJ2M contains nearly two million posts with valid values in the mood and music entries. The content of LJ2M has been enriched with musical content information by using the APIs provided by online music services including 7digital and EchoNest. In sum, the dataset has the following content:

- 1,928,868 blog posts entered by 649,712 online users spontaneously in their daily lives, with verified <article, mood entry, music entry> triplet for each post.
- These posts contain 88,164 unique song titles of 12,201 artists. For most songs, short audio previews (mostly 30 seconds) can be retrieved from 7digital.
- EchoNest track ID by which music content features and metadata can be gathered by the EchoNest API.
- Term-document cooccurrence tables describing the content of each blog article. Moreover, the Linguistic Inquiry and Word Count text analyzer (LIWC) [10] is employed to generate psycholinguistic features.

LJ2M is available at http://mac.citi.sinica.edu.tw/lj/. To the best of our knowledge, LJ2M represents one of the first large-scale datasets characterized by the great diversity of real-life listening contexts in which the emotional state of a user influences the user’s short-term musical preference. From a scientific point of view, LJ2M enables deeper understanding of the interrelationship among the personal, situational, and musical factors of music listening [11]. For applications, LJ2M can be employed for studying context-aware music recommendation [3, 4], emotion-based playlist generation [12], affective human-computer interface [13], and music therapy [14], to name a few.

As an example application, we present research that investigates the interaction between the affective context of the listener and the affective content of music, by using a state-of-the-art music emotion recognition system [12] for analyzing affective content of the audio previews, an affect lexicon called ANEW [15] for emotion visualization, and the psycholinguistic features computed by LIWC for text analysis of the blog articles [10]. The study offers insights into the role of music in mood regulation [16] and demonstrates how LJ2M can be used for studying real-world music listening behaviors.

2. RELATED DATASETS

The million song dataset (MSD) is one of the first large-scale datasets in the MIR community [17]. It contains metadata and audio features of one million tracks. Its ‘EchoNest taste profile subset’ contains some listening records, but there is no contextual or background information about the users.

Yang and Liu have used a smaller LiveJournal dataset LJ40k for a quantitative evaluation, comparing content- and context-based approaches to predict user mood [12]. In contrast, the current work presents a qualitative study on LJ2M.

Some datasets of music-listening records contain user information. For example, Celma [18] compiled two datasets from the last.fm, the first one with time-stamped listening records of 1K users, while the second one with the records of 360K users but no song title and time stamp. Partial gender and age information of the users are available. More recently, Hauger et al. [19] constructed a million-scale music-listening dataset from Twitter by searching for music-related hashtags. This dataset is characterized by the inclusion of geographical information of the user and the timestamps of the tweets, but it does not come with labels of the user’s emotional state.

Many datasets have been built for text-based sentiment analysis in product reviews and social blogs [7, 8]. Recent years have witnessed a growing number of methods and datasets being developed for Twitter. Some works utilized the emoticons to gather emotional state of users [8], while some works resorted to crowdsourcing such as the 2013 SemEval semantic evaluation exercises [9]. Some researchers have also employed LiveJournal for sentiment analysis [7], but little attention if any has been made to investigate the music listening contexts LiveJournal holds.

3. THE LJ2M DATASET

LiveJournal is a social-networking service with 40 million registered users and 1.8 million active users at the end of 2012. In [7], Leshed and Kaye collected 21 million posts from LiveJournal using its RSS feeds for sentiment analysis. The posts they collected were mostly written by users from the United States during years 2000 to 2005. Due to the limitation of the RSS service, there are at most 25 posts for a user. Each post contains an article, a mood entry and a music entry. For the mood entry, users can fill in anything or choose one of the 132 tags pre-defined by LiveJournal. For the music entry, users can also fill in anything. The music entries were

3 http://www.7digital.com/, http://echonest.com/
4 http://www.last.fm/
5 http://en.wikipedia.org/wiki/LiveJournal
6 http://www.livejournal.com/moodlist.bml
not investigated in their study [7]. The raw data were kindly provided by Leshed after personal communication.

We further processed the raw data as follows. For the mood entry, though the users could fill anything as a mood tag, we considered as valid only the 132 mood tags predefined by LiveJournal because users may have filled text which does not correspond to a emotional state.

Table 1 shows the numbers of posts associated with the most and least popular user-mood tags in LJ2M.

<table>
<thead>
<tr>
<th>Most popular user moods</th>
<th>Least popular user moods</th>
</tr>
</thead>
<tbody>
<tr>
<td>tired</td>
<td>infuriated</td>
</tr>
<tr>
<td>happy</td>
<td>surprised</td>
</tr>
<tr>
<td>bored</td>
<td>morose</td>
</tr>
<tr>
<td>amused</td>
<td>embarrassed</td>
</tr>
<tr>
<td>blah</td>
<td>jealous</td>
</tr>
<tr>
<td>content</td>
<td>irate</td>
</tr>
<tr>
<td>cheerful</td>
<td>envious</td>
</tr>
<tr>
<td>calm</td>
<td>recumbent</td>
</tr>
<tr>
<td>contemplative</td>
<td>sympathetic</td>
</tr>
<tr>
<td>sleepy</td>
<td>intimidated</td>
</tr>
</tbody>
</table>

Among them, 64,124 songs (74%) can be found in MSD, so musical metadata and features presented in MSD can be used, such as the EchoNest song-level or segment-level features [17]. If a researcher wants to run other audio feature extraction algorithms, 87,708 songs (99%) have short audio previews available from 7digital.

To avoid privacy issues, the data will be anonymized by replacing the user names with randomly generated IDs. The content of the articles will be provided as lists of word counts with both non-stemmed and stemmed versions. The list of unique terms will be provided, but with less popular terms and name entities removed. On average, an article has 258±330 words. This processing will destroy the structure of the articles and deter further advanced text processing, but in the end we put more weights on the securing of privacy. In addition, we provide the analysis result of the psycholinguistic analysis LIWC [10], which will be described in Section 4.3.

4. EXAMPLE STUDY: THE RELATIONSHIP BETWEEN USER MOOD AND MUSIC EMOTION

As an example application of LJ2M, we present a study that investigates the interaction between the affective context of the listener (referred to as the “user mood”) and the affective content of music (referred to as the “music emotion”). Specifically, we use the mood entries to represent the affective context of users and employ music emotion recognition (MER) techniques [12] to predict the music emotion of the selected music. We note that, while a great deal of research has been made for MER in the MIR community [20, 21], much less efforts have been devoted to studying the influence of user mood on music listening behavior, mostly due to the absence of a dataset that contains user mood information.

According to psychology studies, music is regularly used for mood regulation in our daily life, either for negative mood management, positive mood maintenance, or diversion from boredom [16, 22]. Therefore, we propose to address the following two research questions: 1) what is the correlation between user mood and music emotion in a real-life listening context? and 2) for the particular case of feeling sad, is there any clue from the blog article that is indicative of the preference of sad-sounding (mood-congruent) or happy-sounding (mood-incongruent) music by the blogger?

4.1. Pre-processing: Music Emotion Recognition

The tracks in LJ2M are not labeled with music emotion. It is possible to crawl the web for emotion-related tags. However, the number of mood tags retrieved from web is limited. Therefore, we resorted to state-of-the-art MER techniques and predict music emotion by analyzing the content of the audio previews. The result might contains prediction errors, but it offers more complete information of music emotion.

Specifically, we employed the MER models trained from a last.fm dataset of 31,427 songs [12], which considers a total number of 190 music emotion classes following the taxonomy defined by AllMusic. The problem is formulated as a multi-label classification problem, which entails the training of 190 independent binary-classifiers. The 12-D EchoNest timbre descriptor [17] is adopted as the underlying feature representation, and the RBF kernel of support vector machine [23] is used. The average accuracy of the 190 binary classifiers is 73.9% in AUC (with chance level being 50%), according to cross-validation on the last.fm dataset [12]. With these
MER models, we obtained a 190-D vector for each track in the LJ2M dataset, where each dimension is a numerical value indicating the “affinity” (i.e., degree of association, in $[0,1]$) between the track and a music emotion.

Instead of analyzing the 132 user moods and 190 music emotions, in this study we considered the subsets of 56 user moods and 43 music emotions which can be visualized in the valence–arousal (VA) emotion space according to the affective norm for English words (ANEW) [15]. ANEW is an affect lexicon constructed by psychologists to provide normative ratings for a number of terms in the VA space, where valence indicates affect appraisals (positive/negative) and arousal indicates energy and stimulation level [11–14]. As Fig. 2 displays, by doing so the user moods and music emotions can be mapped to the same low-dimensional space, which facilitates the interpretation of their association.

4.2. Correlations between User Mood & Music Emotion

We analyzed the correlation between each <user mood, music emotion> pair as follows. First, we represented a music emotion tag $i$ by a numerical vector $a_i \in \mathbb{R}^n$ indicating the estimated affinity (in $[0,1]$) by MER for the association between the tag and a post in the data collection, where $n = 1,928,868$ is the total number of posts for the case of LJ2M. Then, we represented a user mood tag $j$ by a binary vector $b_j \in \mathbb{R}^n$ of the same length according to whether that mood tag is labeled with the post by LiveJournal users. Finally, we calculated the Pearson’s correlation coefficient between $a_i$ and $b_j$. The same process was performed for each possible $(i,j)$ pair, and the result was presented in the VA space. As Fig. 3 shows, We drew the top-5 positively correlated music emotions for each user mood, with larger fonts indicating stronger correlation. Moreover, for visualization purpose, Gaussian process regression [12] was applied to interpolate the association over the VA space, with lighter area indicating stronger correlation. Due to space limit, we only show the result of 12 user moods.

The following observations are made. When being in a happy-related mood (i.e., the 1st row of Fig. 3), people tend to listen to happy-sounding music (e.g., ‘happy’ and ‘lively’). Also, when being in a calm mood (2nd row), people prefer music with emotions akin to the mood, such as ‘peaceful.’ However, while being in a sad-related mood (3rd row), the preference diverges: some prefer music with negative emotion, whereas others prefer music with positive emotion. Although similar observations have been made by psychologists in studying the role of music in mood regulation [16, 22], be-

Fig. 3: Result of correlation analysis between user mood (shown in the title of each sub-figure) and music emotion (i.e., tags inside each sub-figure) based on LJ2M.

Fig. 2: Valence and arousal values of 56 user moods (top) and 43 music emotions (down) according to ANEW [15].
ing able to recognize such patterns is encouraging because they are mined from a real-life and possibly noisy dataset LJ2M, and because this data-driven approach is applicable as long as we have relevant data to be analyzed. For instance, we see people prefer angry-sounding music when feeling bored (4th row), a user mood relatively less studied in psychology.

### 4.3. When Feeling Sad: A Psycholinguistic Analysis

From the previous study, we see individual differences in music preference when people feel sad. As the blog articles written by the users possibly contain information about the external context (e.g., activity, weather) and internal context (e.g., mood, personality) of the users, an interesting question is whether we can identify linguistic patterns from the articles that are correlated with such a preference.

For this study, we considered only the posts that are labeled with ‘sad’ user mood and posts with musical tracks whose music emotion can be computed (note that for some tracks the audio previews are not available). This gives rise to \( m = 17,452 \) posts in total. Moreover, instead of using all the 190 music emotions, we considered only the two music emotions ‘happy’ and ‘sad’ for simplicity.

The corpus contains a large number of unique words. For the ease of interpretation, we employed LIWC to convert the word counts of each article into the word ratio with respect to 80 linguistic categories, including function words, verbs, social words, and swear words, to name a few [10]. This way, each article is represented by a 80-dimensional vector.

We computed the correlation between one word category and one music emotion for each word category, music emotion pair. First, a word category \( k \) was represented as a vector \( c_k \in \mathbb{R}^m \) indicating the word ratio of an article with respect to the word category, and a music emotion (‘happy’ or ‘sad,’ referred to as Happy-ME and Sad-ME, respectively) was represented as a vector \( d_i \in \mathbb{R}^m \) of the estimated affinity. Then, we calculated the Pearson’s correlation coefficient between \( c_k \) and \( d_i \) and reported the pairs \((k, i)\) whose correlations are significantly different from zero (\( p \)-value\(<0.001\)) with Student’s t-test. The result is shown in Table 2.

By comparing the result of Happy-ME and Sad-ME, we see that only four LIWC categories are considered correlated with both music emotions, including “assent,” “fuct,” “pronoun,” “ppron.” This shows that the word uses of Happy-ME and Sad-ME are quite different. Moreover, we see the category “assent” is positively correlated with Happy-ME but negatively correlated with Sad-ME, showing that the category is indicative of music preference — people who use more “assent” words (e.g., “yes, ok, lol”) when writing about a sad event tend to listen to happy music when feeling sad.

We also observe relatively high correlation with words in the categories of “sweat,” “you,” “insight,” “cogmech” and “percept” for Sad-ME, and relatively high correlation with words in the categories of “verb,” “i” and “posemo” for Happy-ME.

Tausczik and Pennebaker [10] surveyed and summarized the interrelations between thinking style and word uses as follows: exclusion words (“excl”) are used to make a distinction; conjunction words (“conj”) are used to combine different thoughts together; prepositions (“prep”) indicate that the user is providing more complex and concrete information about the event; words longer than six letters (“sixltr”) suggest more complex language; insight words (“insight”) may reveal that the user is reevaluating the event; whereas the use of “filler” words reveals the uncertainty about a story for the speaker. It happens that almost all of the word indicators of thinking style mentioned by Tausczik and Pennebaker appear in either Happy-ME and Sad-ME, but not in both. This suggests that at least part of the difference of word use in our study is from the difference in thinking style. On the other hand, Rude et al. found positive correlation between depression and the use of “i” words due to the focus on themselves for depressed people [24]. Stirman and Pennebaker also found that poets committing suicide used more “i” in their poems than the poets not committing suicide did [25]. Their works and our observations might have connections, and further investigations are needed to verify it.

### 5. CONCLUSION

In this paper, we have presented LJ2M, one of the largest context-rich datasets for user-centered music information re-
trieval research. The million triplets of <article, mood entry, music entry> allow researchers to study different aspects of real-life music-listening behaviors. We have also described two empirical studies that investigate the relationship among user mood, music emotion, and word use. These studies lead to new insights regarding the role of music in mood regulation and the way thinking style influences music listening behavior. It is hoped that the LJ2M dataset can contribute to the understanding of human behavior related to music and also to the development of user-centric multimedia applications.

6. REFERENCES


